Theoretical Computer Science 410 (2009) 1847–1862

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Polynomial summaries of positive semidefinite kernels

Kilho Shin [a,*], Tetsuji Kuboyama [b]

[a] Carnegie Mellon CyLab Japan, Kobe, Japan

[b] Gakushuin University, Tokyo, Japan

ARTICLE INFO

Keywords:
Machine learning
Kernel method
Polynomial kernel

ABSTRACT

Polynomials have proven to be useful tools to tailor generic kernels to specific applications. Nevertheless, we had only restricted knowledge for selecting *fertile* polynomials which consistently produce positive semidefinite kernels. For example, the well-known polynomial kernel can only take advantage of a very narrow range of polynomials, that is, the univariate polynomials with positive coefficients. This restriction not only hinders intensive exploitation of the flexibility of the kernel method, but also causes misuse of indefinite kernels. Our main theorem significantly relaxes the restriction by asserting that a polynomial consistently produces positive semidefinite kernels, if it has a positive semidefinite coefficient matrix. This sufficient condition is quite natural, and hence, it can be a good characterization of the *fertile* polynomials. In fact, we prove that the converse of the assertion of the theorem also holds true in the case of degree 1. We also prove the effectiveness of our main theorem by showing three corollaries relating to certain applications known in the literature: the first and second corollaries, respectively, give generalizations of the polynomial kernel and the principal-angle (determinant) kernel. The third corollary shows extended and corrected sufficient conditions for the codon-improved kernel and the weighted-degree kernel with shifts to be positive semidefinite.

1. Introduction

To exploit the flexibility of the kernel method, it is critical that sufficiently wide latitude is allowed in selecting kernel functions. In this regard, polynomials have been a useful tool to tailor generic basic kernels (we call them *underlying kernels*) to the context of specific applications. The kernels in the literature that are defined by taking advantage polynomials include the polynomial kernel [2], the principal-angle kernel [7], the determinant kernel [9], the codon-improved kernel [10] and the weighted-degree-with-shift kernel [4]. For example, when a underlying kernel $k(x, y)$ is given, a polynomial kernel $K(x, y)$ with respect to $k(x, y)$ is defined as follows.

$$K(x, y) = \sum_{i=1}^{d} c_i k(x, y)^i.$$

Thus, polynomials have been used as an effective tool to engineer kernels, but only a little was known about intrinsic properties of those polynomials that yield *positive semidefinite* kernels.

Positive semidefiniteness of kernels is a critical premise for many kernel-based learning machines to work properly (*e.g.* SVM [2]). A positive semidefinite kernel has the property that arbitrary Gram matrices with respect to the kernels are positive

* Corresponding address: Kobe Harborland Center Building 17F, 1-3-3 Higashi-Kawasaki, Chuo, Kobe 6500044, Japan. Tel.: +81 78 360 5311; fax: +81 78 360 1617.

E-mail addresses: yshin@cmuj.jp (K. Shin), kuboyama@tk.cc.gakushuin.ac.jp (T. Kuboyama).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2009.01.014

semidefinite (*i.e.* they are symmetric and don't have negative eigenvalues), where the Gram matrix $G(x_1, \ldots, x_n : K)$ for an arbitrary finite set $\{x_1, \ldots, x_n\}$ in the space χ of data points is defined as follows.

$$G(x_1, \ldots, x_n : K) = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \ldots & K(x_1, x_n) \\ K(x_2, x_1) & K(x_2, x_2) & \ldots & K(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n, x_1) & K(x_n, x_2) & \ldots & K(x_n, x_n) \end{bmatrix}.$$

If χ is a finite set, the positive semidefiniteness is equivalent to the property that there exists a mapping (*feature decomposition*) $\Phi : \chi \longrightarrow \mathbb{R}^N$ such that $K(x, y) = \Phi(x)^\top \Phi(y)$. Positive semidefinite kernels are also known as *reproducing* kernels and *Mercer's* kernels.

As for polynomial kernels, we have a good condition with respect to the coefficients c_i to make them positive semidefinite. In fact, polynomial kernels are always positive semidefinite for $c_i \geq 0$ and for positive semidefinite $k(x, y)$. This condition is particularly useful, since we have latitude to optimize the coefficients so that classification performance becomes the best.

This condition, however, has only a narrow range of application. For example, it cannot explain the positive semidefiniteness of the determinant kernel, since the polynomial for the determinant kernel includes negative coefficients (see the second row of Table 2). In [9], the positive semidefiniteness of the determinant kernel is proven in a straightforward and case-specific manner. In fact, a feature decomposition was directly defined, taking advantage of some discriminative property of the determinant kernel and Binet–Cauchy formula as follows. Let data points x and y be represented by $D' \times D$ real matrices X and Y for $D' \geq D$. Then, the determinant kernel is defined by $K(x, y) = \det\left[X^\top Y\right]$. By Binet–Cauchy formula, the following equality holds with the D-th order minors of X and Y.

$$\det\left[X^\top Y\right] = \sum_{1 \leq k_1 < \cdots < k_D \leq D'} X\begin{pmatrix} k_1 & k_2 & \ldots & k_D \\ 1 & 2 & \ldots & D \end{pmatrix} Y\begin{pmatrix} k_1 & k_2 & \ldots & k_D \\ 1 & 2 & \ldots & D \end{pmatrix}.$$

When $x_{i,j}$ denotes the (i, j)-element of X, the D-th order minor of X is

$$X\begin{pmatrix} k_1 & k_2 & \ldots & k_D \\ 1 & 2 & \ldots & D \end{pmatrix} = \det \begin{bmatrix} x_{k_1,1} & x_{k_1,2} & \ldots & x_{k_1,D} \\ x_{k_2,1} & x_{k_2,2} & \ldots & x_{k_2,D} \\ \vdots & & & \vdots \\ x_{k_D,1} & x_{k_D,2} & \ldots & x_{k_D,D} \end{bmatrix}.$$

Thus, when we map x and y into a feature space so that their coordinates are exactly identical to the D-th minors of X and Y, $K(x, y)$ is nothing other than the inner product between the images of x and y in the feature space.

Motivated by the recognition of this contrast, we will address, in this paper, the problem that we don't have good generic conditions with respect to the coefficients of polynomials such that:

- we can use them to discriminate between the *fertile polynomials* that always yield positive semidefinite kernels and those that don't necessarily;
- they have a sufficiently wide range of application when used for the aforementioned discrimination purpose.

In fact, this paper presents a sufficient condition for polynomials to be *fertile* (Theorem 9, Section 5). The form of the condition is very general, and we may have a wide latitude in selecting the coefficients of polynomials without harming the positive semidefiniteness of the resulting kernels. Moreover, the condition turns out to be a *necessary* condition in the case of degree 1 (Section 7). To show that the range of application of the condition is wide, we will see that the positive semidefiniteness of the kernels referenced in the above can be verified according to the condition (Section 8).

2. Preliminaries and notations

Throughout this paper, we assume that a matrix has only real elements. When the transpose of a matrix A is denoted by ${}^\top A$, a symmetric A satisfies ${}^\top A = A$, and an orthogonal A satisfies ${}^\top A = A^{-1}$. The trace of A is defined as the sum of its diagonal elements, and is denoted by $\mathrm{tr}(A)$.

As mentioned in the previous section, a positive semidefinite kernel $K : \chi \times \chi \longrightarrow \mathbb{R}$ is defined so that, for arbitrary $x_1, \ldots, x_n \in \chi$, the Gram matrix $G(x_1, \ldots, x_n : K)$ is positive semidefinite.

To define the positive semidefinite matrix and the positive definite matrix, we note Proposition 1 and Proposition 2, which present important properties of the symmetric matrices.

Proposition 1. *For an n-dimensional symmetric real matrix A, the following are equivalent to each other.*

(1) $(c_1, \ldots, c_n)A(c_1, \ldots, c_n)^\top \geq 0$ *for arbitrary* $(c_1, \ldots, c_n) \in \mathbb{R}^n$.
(2) *A has only non-negative real eigenvalues.*
(3) *There exists an n-dimensional orthogonal matrix P such that* $P^\top AP$ *is a diagonal matrix with non-negative elements.*
(4) $A = B^\top B$ *for some n-dimensional real matrix B.*
(5) $A = B^\top B$ *for some* $m \times n$ *real matrix B.*

Proposition 2. *For an n-dimensional symmetric real matrix A, the following are equivalent to each other.*

(1) $(c_1, \ldots, c_n)A(c_1, \ldots, c_n)^\top > 0$ *for arbitrary* $(c_1, \ldots, c_n) \in \mathbb{R}^n \setminus \{(0, \ldots, 0)\}$.
(2) *A has only positive real eigenvalues.*
(3) *There exists an n-dimensional orthogonal matrix P such that* $P^\top AP$ *is a diagonal matrix with positive diagonal elements.*
(4) $A = B^\top B$ *for some n-dimensional real matrix B with* $\det B \neq 0$.
(5) $A = B^\top B$ *for some* $m \times n$ *real matrix B of rank m.*

Now, we define the positive semidefinite matrix and the positive definite matrix as follows.

Definition 3. A real matrix A is called *positive semidefinite (positive definite)*, if, and only if, A is symmetric and one of, hence all of, the conditions of Proposition 1 *(Proposition 2)* hold for A.

When $A = [a_{ij}]$ is an n-dimensional matrix, $A[\alpha_1, \ldots, \alpha_\ell]$ denotes the ℓ-dimensional matrix whose (i, j)-element is identical to $a_{\alpha_i\alpha_j}$, where $\{\alpha_1, \ldots, \alpha_\ell\}$ is a subset of $\{1, \ldots, n\}$. Then, we have useful characterization of the positive definite matrix and the positive semidefinite matrix as stated below.

Proposition 4 (*Theorem 1.16, [1]*). *An n-dimensional symmetric matrix A is positive definite, if, and only if,* $\det A[1, \ldots, \ell] > 0$ *for any* ℓ *such that* $1 \leq \ell \leq n$.

Proof. The "only if" part is shown as follows. Since A is positive definite, there exists an $m \times n$ matrix B of rank n such that $A = B^\top B$. When $\vec{b}_i$ denotes the i-th column vector of B, $A[1, \ldots, \ell] = \left[\vec{b}_1, \ldots, \vec{b}_\ell\right]^\top \left[\vec{b}_1, \ldots, \vec{b}_\ell\right]$ implies our assertion.

The "if" part can be proven by induction on n. Since the assertion is trivial for $n = 1$, we assume $n > 1$. By the hypothesis of induction, there exists an $(n-1)$-dimensional orthogonal matrix P such that $P^\top A[1, \ldots, n-1]P$ is diagonal with positive diagonal elements. Since proving that A is positive definite is equivalent to proving that, so is $\begin{bmatrix} P^\top & \vec{0}^\top \\ \vec{0} & 1 \end{bmatrix} A \begin{bmatrix} P & \vec{0}^\top \\ \vec{0} & 1 \end{bmatrix}$, we may assume that $A[1, \ldots, n-1]$ is diagonal without loss of generality.

$$\det A = \left(\prod_{i=1}^{n-1} a_{ii}\right) \cdot \left(a_{nn} - \sum_{i=1}^{n-1} \frac{a_{in}^2}{a_{ii}}\right) > 0$$

$$(c_1, \ldots, c_n)A(c_1, \ldots, c_n)^\top = \sum_{i=1}^{n-1} a_{ii} \left(c_i + \frac{a_{in}}{a_{ii}} c_n\right)^2 + \left(a_{nn} - \sum_{i=1}^{n-1} \frac{a_{in}^2}{a_{ii}}\right) c_n^2.$$

Our assertion immediately follows from these formulas. □

Proposition 5 (*Theorem 1.17, [1]*). *An n-dimensional symmetric matrix A is positive semidefinite, if, and only if,* $\det A[\alpha_1, \ldots, \alpha_\ell] \geq 0$ *for an arbitrary subset* $\{\alpha_1, \ldots, \alpha_\ell\} \subsetneqq \{1, \ldots, n\}$.

Proof. The "only if" part is proven in the same way as Proposition 4. We prove the "if" part tracing the proof presented in [1]. Let Δ_m be the m-dimensional identity matrix. Then, we have $\det(A[1, \ldots, m] + \epsilon\Delta_m) = \sum_{p=0}^{m} d_{m,p}\epsilon^p$, where $d_{m,m} = 1$ and $d_{m,p} = \sum_{1\leq\alpha_1<\cdots<\alpha_{m-p}\leq m} \det A[\alpha_1, \ldots, \alpha_{m-p}]$. In particular,

$$\det(A[1, \ldots, m] + \epsilon\Delta_m) \geq \epsilon^m > 0$$

holds for $\epsilon > 0$, and therefore, $A + \epsilon\Delta_n$ is positive definite by Proposition 4. The assertion follows from $(c_1, \ldots, c_n)A(c_1, \ldots, c_n)^\top = \lim_{\epsilon\to+0}(c_1, \ldots, c_n)(A + \epsilon\Delta_n)(c_1, \ldots, c_n)^\top$. □

3. Problem identification and our contributions

In this section, we first review a few instances of the kernels that are defined taking advantage of polynomials, and see that they were defined in case-specific manners (3.1). Then, we identify the problem that we will focus on in this paper (3.2), and then summarize our solutions to the problem (3.3).

3.1. A review of polynomial-based composition of kernels

To start with, we will visit 5 kernels in the literature, all of which are defined by means of the polynomial-based composition.

3.1.1. The polynomial (Poly) kernel:

The *polynomial kernel* is given in the form of $(k(x, y) + c)^d$ for an underlying kernel $k(x, y)$. It is known that a polynomial kernel is positive semidefinite, if c is non-negative and $k(x, y)$ is positive semidefinite (*e.g.* [2]). The polynomial kernel has proven useful for two main reasons — (1) a separating *hypersurface*[1] in the feature space of the underlying kernel is mapped to a separating *hyperplane* in a higher-dimensional feature space; (2) the polynomial kernel reflects the correlation of tuples of features of the underlying kernel [10]. Polynomial kernels can be generalized to the form of $K(x, y) = \sum_{i=0}^{d} c_i k(x, y)^i$ with arbitrary $c_i \geq 0$ without harming the positive semidefiniteness.

3.1.2. The determinant (Det) kernel and the principal-angle (PA) kernel:

Zhou [9] introduced the determinant kernel for matrix-type data points — when data points x and y are respectively represented by matrices $X = [x_{ij}]$ and $Y = [y_{ij}]$ for $i = 1, \ldots, D'$ and $j = 1, \ldots, D$, the determinant kernel for x and y is defined as follows.

$$K(x, y) = \det[X^{\mathsf{T}} Y].$$

Also, under the same setting as Zhou [9], Wolf et al. [7] showed that principal angles $(\theta_1, \ldots, \theta_D)$ of the column spaces of the matrices X and Y can be efficiently computed using the *kernel trick*, and introduced the principal-angle kernel (1), where Q_X and Q_Y are the matrices obtained by the QR decomposition of X and Y.

$$K(x, y) = \left(\det[{Q_X}^{\mathsf{T}} Q_Y]\right)^2 = \prod_{i=1}^{D} \cos^2 \theta_i. \tag{1}$$

It is known that the principal angles can be effective measures for similarity between such data points represented by matrices (*e.g.* [8]).

When we denote the i-th column vector of X for the determinant kernel and Q_X for the principal-angle kernel by $x^{(i)}$, the data point x is equivalently represented as the sequence $(x^{(1)}, \ldots, x^{(D)})$. Then, the positive semidefiniteness of the determinant kernel and the principal-angle kernel is reduced to that of the kernel $K(x, y)$ defined as follows.

$$k(x^{(i)}, y^{(j)}) = {x^{(i)}}^{\mathsf{T}} y^{(j)}$$

$$K(x, y) = \det \begin{bmatrix} k(x^{(1)}, y^{(1)}) & \ldots & k(x^{(1)}, y^{(D)}) \\ \vdots & \ddots & \vdots \\ k(x^{(D)}, y^{(1)}) & \ldots & k(x^{(D)}, y^{(D)}) \end{bmatrix}. \tag{2}$$

Indeed, the kernel (2) is an example of the kernels composed by applying polynomials to positive semidefinite underlying kernels, since we have

$$K(x, y) = \sum_{\sigma \in \mathfrak{S}_D} \operatorname{sgn}(\sigma) \prod_{i=1}^{D} k(x^{(i)}, y^{(\sigma(i))}).$$

In the original papers, the positive semidefiniteness of the kernel (2) was proven by Binet–Cauchy theorem [7,9,6] (see also Section 1). In Section 8.2, we will prove the property as a corollary to our main theorem.

3.1.3. The codon-improved (CI) kernel and the weighted-degree-with-shift (WDwS) kernels:

The codon-improved kernel [10] and its generalization, the weighted-degree-with-shift kernel [4] are similar to the spectrum kernel [3] in that they count the matching substrings between a pair of strings, but are different in that matches are weighted according to their positional information. Although these kernels are defined using polynomials, different from the examples seen so far, their positive semidefiniteness is not proven in a straightforward manner. In fact, careful selection of the coefficients of the polynomials is required to maintain the positive semidefiniteness, and both [10] and [4] made mistakes in this regard.

For example, the codon-improved kernel is designed so as to exploit the *a priori* knowledge "*a coding sequence (CDS) shifted by three nucleotides still looks like CDS*" [10]. In fact, in addition to the matches of substrings that start at the same position, it counts those whose starting positions differ exactly by 3. A precise definition is given as follows. For sequences of nucleotides x and y, we let x_p (resp. y_p) denote the nucleotide at position p in x (resp. y). Then, $k_p(x, y)$ is defined as follows.

$$k_p(x, y) = \sum_{j=-\ell}^{\ell} w_{|j|} \delta(x_{p+j}, y_{p+j}). \tag{3}$$

In (3), $w_{|j|}$ are non-negative weights, and $\delta(x_{p+j}, y_{p+j})$ is Kronecker's delta: $\delta(x_{p+j}, y_{p+j})$ is 1, if x_{p+j} and y_{p+j} represent the same nucleotide, and it is 0, otherwise. When T_3 denotes the 3-shift operator that chops off the leading 3 nucleotides, the

[1] In this paper, by a hypersurface, we mean a subspace of a Euclidean space defined by an algebraic equation whose degree is higher than 1.

window score $\text{win}_p(x, y)$ at position p and the codon-improved kernel $K(x, y)$ are respectively defined by the formulas (4) and (5), where $\bar{w}_3$ is another non-negative weight and L is the common length of x and y (*i.e.* $L = |x| = |y|$).

$$\text{win}_p(x, y) = \left[k_p(x, y) + \bar{w}_3\left\{k_p(T_3x, y) + k_p(x, T_3y)\right\}\right]^{d_1} \tag{4}$$

$$K(x, y) = \left(\sum_{p=\ell+1}^{L-\ell} \text{win}_p(x, y)\right)^{d_2}. \tag{5}$$

Although Zien et al. [10] claimed that the codon-improved kernels are unconditionally positive semidefinite, the fact is that the weights should be chosen appropriately. We will illustrate this by a simplified example. We assume $w_j = 1$ ($j = -\ell, \ldots, \ell$), $d_1 = d_2 = 1$, $\ell = 3q$, $p = 3q + 1$ and $L = 6q + 1$, and define x and y as follows.

$$x = \underbrace{\texttt{ATGCGT ATGCGT}\ldots\texttt{ATGCGT}}_{6q}\texttt{A}$$

$$y = \underbrace{\texttt{CTGAGT CTGAGT}\ldots\texttt{CTGAGT}}_{6q}\texttt{C}$$

Then, we have $K(x, x) = K(y, y) = 6q + 1$ and $K(x, y) = 4q + 4\bar{w}_3q$, and therefore, the determinant of the Gram matrix for x and y is not necessary non-negative.

$$\begin{vmatrix} K(x,x) & K(x,y) \\ K(y,x) & K(y,y) \end{vmatrix} = \begin{vmatrix} 6q+1 & 4q+4\bar{w}_3q \\ 4q+4\bar{w}_3q & 6q+1 \end{vmatrix} = (10q + 4\bar{w}_3q + 1)(2q - 4\bar{w}_3q + 1).$$

In particular, $\bar{w}_3 \leq \frac{1}{2}$ is necessary for $K(x, y)$ to be positive semidefinite for any q.

Ratsch et al. [4] proceeded along the same line as Zien et al., and introduced the weighted degree kernel with shifts. Different from the codon-improved kernel, the weighted degree kernel with shifts includes shift weights $\bar{w}_s$ for multiple s. First, define the underlying kernel $k(x, y)$ as follows.

$$k(x, y) = \sum_{n=1}^{d} \beta_n \sum_{i=1}^{L-n+1} w_i \delta(x[i, n], y[i, n]), \quad \beta_n \geq 0, \quad w_i \geq 0.$$

The symbol L denotes the upper bound of the lengths of x and y, and $x[i, n]$ does the n-length contiguous substring $x_i x_{i+1} \ldots x_{i+n-1}$ of x starting at position i. For $i + n - 1 > \min\{|x|, |y|\}$, we define $\delta(x[i, n], y[i, n]) = 0$. Finally, for the s-shift operator T_s, the weighted degree kernel with shifts is defined as follows.[2]

$$K(x, y) = k(x, y) + \sum_{s=1}^{S} \bar{w}_s\{k(T_sx, y) + k(x, T_sy)\}. \tag{6}$$

In [4], the positive semidefiniteness of $K(x, y)$ was investigated as follows. Since $k(x, y)$ is positive semidefinite, so is $k(x, y) + k(T_sx, y) + k(x, T_sy) + k(T_sx, T_sy)$. If w_i's take the same value w for all i, as is assumed in [4] with no declaration, $k(x, y) - k(T_sx, T_sy)$ is identical to $\sum_{n=1}^{d} \beta_n w \sum_{i=1}^{s} \delta(x[i, n], y[i, n])$. Thus, $k(x, y) - k(T_sx, T_sy)$ is positive semidefinite, and therefore, so is $2k(x, y) + k(T_sx, y) + k(x, T_sy)$. On the other hand, $K(x, y)$ is evaluated as follows.

$$K(x, y) = \left(1 - 2\sum_{s=1}^{S} \bar{w}_s\right) k(x, y) + \sum_{s=1}^{S} \bar{w}_s \left(2k(x, y) + k(T_sx, y) + k(x, T_sy)\right).$$

Therefore, we can conclude that $\sum_{s=1}^{S} \bar{w}_s \leq \frac{1}{2}$ is a sufficient condition for $K(x, y)$ to be positive semidefinite, since $\bar{w}_s$'s are non-negative.

We can relax the constraint of $w_1 = w_2 = \cdots = w_L$ in the aforesaid discussion to $w_1 \leq w_2 \leq \cdots \leq w_L$, since the following formula implies the same key property that $2k(x, y) + k(T_sx, y) + k(x, T_sy)$ is positive semidefinite.

$$k(x, y) - k(T_sx, T_sy) = \sum_{n=1}^{d} \beta_n \left(\sum_{i=1}^{s} w_i \delta(x[i, n], y[i, n]) + \sum_{i=s+1}^{L-n+1} (w_i - w_{i-s}) \delta(x[i, n], y[i, n])\right).$$

Our main theorem generalizes these results. In fact, we will show in Section 8.3 that the following condition can take over the condition $\sum_{s=1}^{S} \bar{w}_s \leq \frac{1}{2}$.

$$\sum_{s=1}^{S} \frac{\bar{w}_s}{b_s} \leq 1, \quad b_s = \min\left\{ \frac{w_j}{w_j + w_{j-s}} \,\middle|\, j = s+1, s+2, s+3, \ldots, L \right\}.$$

If $w_1 \leq w_2 \leq \cdots \leq w_L$, we have $b_s \geq \frac{1}{2}$.

2 Although S varies according to n in [4], we assume that S is a constant just for simplicity.

Table 1
Polynomials in examples.

Kernels	Variables	Polynomial
Poly	ξ	$\sum_{j=1}^{d} c_j \xi^i, \quad c_j \geq 0$
PA, Det	$\{\xi_{i,j}\}_{i,j=1}^{D}$	$\sum_{\sigma \in \mathfrak{S}_D} \operatorname{sgn}(\sigma) \prod_{i=1}^{D} \xi_{i,\sigma(i)}$
win_p for CI	$\{\xi_{i,j}\}_{i,j=1}^{L+3}$	$\left[\sum_{j=-\ell}^{\ell} w_{\lvert j\rvert} \left\{\xi_{p+j,p+j} + \bar{w}_3(\xi_{p+j,p+j+3} + \xi_{p+j+3,p+j})\right\}\right]^{d_1}$
WDwS with single n	$\{\xi_{i,j}\}_{i,j=1}^{L}$	$\sum_{i=1}^{L-n+1} w_i \xi_{i,i} + \sum_{s=1}^{S} \bar{w}_s \left(\sum_{i=1}^{L-n-s+1} w_i \left(\xi_{i,i+s} + \xi_{i+s,i}\right)\right)$

3.2. Problem identification

Table 1 gives the list of the polynomials used in the examples mentioned in Section 3.1. Of the listed polynomials, that for the polynomial kernel is in the most generic form, but its range of application is nevertheless restricted for two reasons: (1) it is simply univariate; (2) it cannot include negative coefficients. In fact, the kernels referenced in Sections 3.1.2 and 3.1.3 take advantage of multivariate polynomials, and in addition, the principal-angle and determinant kernels include negative coefficients.

Eventually, for the purpose of discriminating between the *fertile polynomials* that yield positive semidefinite kernels and those that don't necessarily, no widely applicable condition was known. This not only would restrict the range of polynomials which we can take advantage of to *engineer* new kernels, but also could cause misuse of indefinite kernels.

The present paper addresses this problem, and, in fact, presents a strong sufficient condition for the *fertile polynomials*.

3.3. Our contributions

Below, the contributions of the present paper are summarized.

- 4 settings are known for deriving polynomial-based kernels from underlying kernels (4.1). In Section 4, we first show that one of the settings is truly more expressive than the others — any kernel derived in the settings other than the most expressive setting can be converted into a kernel derived in the setting, but the converse doesn't hold. Then, we define the *polynomial summary* of kernels under the setting (Definition 8, 4.2). In particular, the polynomial summary with respect to a polynomial p and an underlying kernel $k(x, y)$ is called the *p-summary* of $k(x, y)$.
- Our main theorem (Theorem 9, 5.2) presents a sufficient condition on a polynomial p so that the p-summaries result in positive semidefinite kernels regardless of the choice of the positive semidefinite underlying kernel. A proof to the theorem is given in Section 6. Furthermore, in the case of degree 1, the condition is also a necessary one, in the sense that, if a polynomial p of degree 1 doesn't meet the condition, there exists a positive semidefinite underlying kernel $k(x, y)$ such that the p-summary of $k(x, y)$ is not positive semidefinite (Section 7).
- We introduce three corollaries to Theorem 9. The first two corollaries, respectively, generalize the polynomial kernel and the principal-angle (determinant) kernel (8.1 and 8.2). The third one presents a corrected sufficient condition for the codon-improved kernel and the weighted-degree-with-shift kernel to be positive semidefinite (8.3).

4. Polynomial summaries

In this section, we first pursue the *most expressive* setting for the polynomial-based composition of kernels (4.1), and then define the *polynomial summary* under the setting (4.2).

4.1. Relation among the known settings

In the literature, we can see 4 settings for the polynomial-based kernels according to the answers to the following questions (Type A to D, Table 2).

Q.1 Is the domain of the resulting kernel $K(x, y)$ simply identical with that of the underlying kernel? Or, is it a non-trivial Cartesian product (direct product) of the domain(s) of the underlying kernel(s)?
Q.2 Is only a single underlying kernel to be used, or are multiple underlying kernels to be used?

Type A (Domain: Simple, Underlying Kernel: Single). This type faithfully represents the polynomial kernel. A univariate polynomial $p(\xi) = \sum_{i=0}^{d} c_i \xi^i$ is applied to a single underlying kernel $k : \chi \times \chi \to \mathbb{R}$. The resulting kernel $K(x, y)$ is simply defined as $p(k(x, y)) = \sum_{i=0}^{d} c_i k(x, y)^i$, which is positive semidefinite, if $c_i \geq 0$ for all i.

Type B (Domain: Product, Underlying Kernel: Single). A multivariate polynomial p in the D^2 variables ξ_{ij} for $i, j = 1, \ldots, D$ is applied to a single underlying kernel $k : \chi_* \times \chi_* \to \mathbb{R}$. The domain χ is defined as χ_*^D, and $K((x_1, \ldots, x_D), (y_1, \ldots, y_D))$ is obtained by substituting $k(x_i, y_j)$ for ξ_{ij}. The aforesaid examples except the polynomial kernel belong to this type.

Table 2
Types of polynomial-based kernels.

Type	A	B
Domain of K	χ	χ_*^D
Underlying kernel(s)	$k : \chi \times \chi \to \mathbb{R}$	$k : \chi_* \times \chi_* \to \mathbb{R}$
Polynomial	$p(\xi)$	$p(\xi_{11}, \dots, \xi_{ij}, \dots, \xi_{DD})$
Substitution	$\xi = k(x, y)$	$\xi_{ij} = k(x_i, y_j)$
Example	Poly	PA, Det, CI, WDwS
Type	**C**	**D**
Domain of K	χ	$\chi_1 \times \cdots \times \chi_D$
Underlying kernel(s)	$\{k_d : \chi \times \chi \to \mathbb{R}\}_{d=1,\dots,D}$	$\{k'_d : \chi_d \times \chi_d \to \mathbb{R}\}_{d=1,\dots D}$
Polynomial	$p(\xi_1, \dots, \xi_D)$	$p(\xi_1, \dots, \xi_D)$
Substitution	$\xi_d = k_d(x, y)$	$\xi_d = k'_d(x_d, y_d)$
Example	$k_1(x, y) + k_2(x, y)$	$k'_1(x_1, y_1) + k'_2(x_2, y_2)$
	$k_1(x, y) \cdot k_2(x, y)$	$k'_1(x_1, y_1) \cdot k'_2(x_2, y_2)$

Type C (Domain: Simple, Underlying Kernel: Multiple). A multivariate polynomial p in the D variables ξ_d for $d = 1, \dots, D$ is applied to multiple underlying kernels $k_d : \chi \times \chi \to \mathbb{R}$. $K(x, y)$ is obtained by substituting $k_d(x, y)$ for ξ_d. For example, $k_1(x, y) + k_2(x, y)$ and $k_1(x, y) \cdot k_2(x, y)$ are known to be positive semidefinite, if so are k_1 and k_2 (*e.g.* [2, Proposition 3.12]).

Type D (Domain: Product, Underlying Kernel: Multiple). A multivariate polynomial p in the D variables ξ_d for $d = 1, \dots, D$ is applied to multiple underlying kernels $k'_d : \chi_d \times \chi_d \to \mathbb{R}$. The domain χ is defined as $\chi_1 \times \cdots \times \chi_D$, and $K((x_1, \dots, x_D), (y_1, \dots, y_D))$ is obtained by substituting $k'_d(x_d, y_d)$ for ξ_d. For example, when $k'_i : \chi_i \times \chi_i \to \mathbb{R}$ are positive semidefinite for $i = 1, 2$, $k'_1(x_1, y_1) + k'_2(x_2, y_2)$ and $k'_1(x_1, y_1) \cdot k'_2(x_2, y_2)$ define positive semidefinite kernels over $\chi_1 \times \chi_2$.

Apparently, Type A is the special case of the other types where $D = 1$. Let $\mathcal{A}$ denote the set of the *fertile* polynomials that always yield positive semidefinite kernels under the setting of Type A. $\mathcal{A} \subsetneq \mathbb{R}[\xi]$ holds. Further, when $\mathcal{T}$ is one of $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$, which respectively correspond to Type B, Type C and Type D, let $\mathcal{T}_D$ denote the set of the fertile polynomials with respect to the parameter D for the respect type corresponding to $\mathcal{T}$.

$$\mathcal{B}_D \subsetneq \mathbb{R}[\xi_{11}, \dots, \xi_{ij}, \dots, \xi_{DD}], \quad \mathcal{C}_D \subsetneq \mathbb{R}[\xi_1, \dots, \xi_D], \quad \mathcal{D}_D \subsetneq \mathbb{R}[\xi_1, \dots, \xi_D].$$

Then, we have the following relation.

$$\mathcal{A} = \mathcal{B}_1 = \mathcal{C}_1 = \mathcal{D}_1. \tag{7}$$

Also, a kernel of Type D can be regarded as of Type C, when we let $\chi = \chi_1 \times \cdots \times \chi_D$ and $k_d((x_1, \dots, x_D)) = k'_d(x_d)$. Furthermore, when p only includes the variables ξ_{ii}, and when we identify the variable ξ_{ii} with ξ_i, a kernel of Type B for such p can be regarded as of Type D, when we let $\chi_1 = \cdots = \chi_D = \chi_*$ and $k'_1 = \cdots = k'_D = k$. These properties imply the following inclusion relation.

$$\mathcal{B}_D \cap \mathbb{R}[\xi_{11}, \dots, \xi_{ii}, \dots, \xi_{DD}] \supseteq \mathcal{D}_D \supseteq \mathcal{C}_D$$

Lemma 6 asserts that a kernel of Type C can be regarded as of Type B.

Lemma 6. *For an arbitrary family of positive semidefinite kernels $\{k_d : \chi \times \chi \to \mathbb{R}\}_{d=1,\dots,D}$, there exist a set χ_*, a positive semidefinite kernel $k : \chi_* \times \chi_* \to \mathbb{R}$ and an inclusion mapping $i : \chi \to \chi_*^D$ such that $k_d(x, y) = k(i_d(x), i_d(y))$ where $i(x) = (i_1(x), \dots, i_D(x))$ and $i(y) = (i_1(y), \dots, i_D(y))$.*

Proof. Let χ_* be $\chi \times \{1, 2, \dots, D\}$. When $k : \chi_* \times \chi_* \to \mathbb{R}$ is defined so that the Eq. (8) holds, it is obvious that k is positive semidefinite.

$$k((x, a), (y, b)) = \begin{cases} k_d(x, y) & \text{if } a = b = d, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We have only to define $i : \chi \to \chi_*^D$ by $i(x) = ((x, 1), \dots, (x, D))$. $\square$

We identify the variable ξ_i with ξ_{ii} for $i = 1, \dots, D$. For a polynomial $p(\xi_1, \dots, \xi_D) = p(\xi_{11}, \dots, \xi_{ii}, \dots, \xi_{DD})$, Eq. (9) follows from Lemma 6.

$$p(k_1(x, y), \dots, k_D(x, y)) = p(k(i_1(x), i_1(y)), \dots, k(i_D(x), i_D(y))). \tag{9}$$

If a polynomial p is such that $p(k(x_1, y_1), \dots, k(x_D, y_D))$ is positive semidefinite for an arbitrary positive semidefinite k, so is $p(k_1(x, y), \dots, k_D(x, y))$ for an arbitrary positive semidefinite k_d. Therefore, we have the following corollary to Lemma 6.

Corollary 7. *The following relation holds for any positive integer D.*

$$\mathcal{B}_D \cap \mathbb{R}[\xi_{11}, \dots, \xi_{ii}, \dots, \xi_{DD}] = \mathcal{C}_D = \mathcal{D}_D. \tag{10}$$

By the formulas (7) and (10), we conclude that Type B is more expressive than the other types.

4.2. Definition of polynomial summaries

Based on the consideration in the previous subsection, we define the polynomial summary assuming the setting of Type B.

Definition 8. Let $p(\xi_{11}, \xi_{12}, \dots, \xi_{ij}, \dots, \xi_{DD})$ be a real polynomial in the D^2 variables of $\{\xi_{ij} \mid i, j = 1, \dots, D\}$. The *p-summary* of an underlying kernel $k : \chi \times \chi \longrightarrow \mathbb{R}$ is the kernel $p[k] : \chi^D \times \chi^D \longrightarrow \mathbb{R}$ defined as below.

$$p[k]((x_1, \dots, x_D), (y_1, \dots, y_D)) = p(k(x_1, y_1), k(x_1, y_2), \dots, k(x_i, y_j), \dots, k(x_D, y_D)).$$

Example 1. The kernel (2) is a polynomial summary with respect to the polynomial $\Phi_{D,0}$ defined below.

$$\Phi_{D,0}(\xi_{11}, \dots, \xi_{DD}) = \sum_{\sigma \in \mathfrak{S}_D} \operatorname{sgn}(\sigma) \prod_{i=1}^{D} \xi_{i\sigma(i)}.$$

In [6,9], it is shown that, if the underlying kernel k is positive semidefinite, the $\Phi_{D,0}$-summary of k is positive semidefinite. Hence, we have $\Phi_{D,0} \in \mathcal{B}_D$.

Example 2. Define f_p as follows.

$$f_p(\xi_{11}, \dots, \xi_{DD}) = \left[\sum_{j=-\ell}^{\ell} w_j \left\{ \xi_{p+j,p+j} + \bar{w}_3(\xi_{p+j,p+j+3} + \xi_{p+j+3,p+j}) \right\} \right]^{d_1}.$$

Then, the window score $\mathrm{win}_p(x, y)$ of Eq. (4) is the f_p-summary of Kronecker's delta $\delta(\xi, \eta)$ defined over the alphabet *{A, T, G, C}*. Although $\delta(\xi, \eta)$ is positive semidefinite, it is necessary to choose appropriate w_j and $\bar{w}_3$ to make the resulting f_p-summary positive semidefinite (3.1.3 and 8.3).

As seen in Example 2, even if the underlying kernel k is positive semidefinite, the p-summary of k may or may not be positive semidefinite, dependent on the choice of the polynomials p. In other words, $\mathcal{B}_D$ is a proper subset of $\mathbb{R}[\xi_{11}, \dots, \xi_{ij}, \dots, \xi_{DD}]$.

$$\mathcal{B}_D \subsetneqq \mathbb{R}[\xi_{11}, \dots, \xi_{ij}, \dots, \xi_{DD}].$$

Thus, the following question naturally arises.

When do the polynomial summaries of polynomials p become positive semidefinite kernels?

Theorem 9 answers this question by presenting a sufficient condition for such polynomials.

5. The main theorem

Our main theorem asserts that, if a polynomial p has a positive semidefinite coefficient matrix, the p-summary $p[k]$ of a positive semidefinite underlying kernel k is always positive semidefinite. In this section, we will first give the definition of the coefficient matrix (5.1), and then, will provide the statement of our main theorem (5.2). After introducing a certain lemma that plays a key role in proving the theorem (5.3), we will sketch a plot of the proof of the theorem that will be given in Section 6 (5.4).

5.1. Coefficient matrices of polynomials

Assume that a polynomial p of degree d in the D^2 variables $\xi_{11}, \dots, \xi_{ij}, \dots, \xi_{DD}$ is given the following representation, where Δ and $\vec{\emptyset}$ denote the set $\{1, \dots, D\}$ and the empty sequence, respectively.

$$p = c_{\vec{\emptyset},\vec{\emptyset}} + \sum_{\delta=1}^{d} \sum_{(k_1,\dots,k_\delta)\in\Delta^\delta} \sum_{(l_1,\dots,l_\delta)\in\Delta^\delta} c_{(k_1,\dots,k_\delta),(l_1,\dots,l_\delta)} \cdot \xi_{k_1 l_1} \xi_{k_2 l_2} \cdots \xi_{k_\delta l_\delta}.$$

Then, we define a $\frac{D^{d+1}-1}{D-1}$-dimensional square matrix C.

- The columns and rows of C are indexed by the vector $\vec{i} \in \Delta^{[0..d]} = \bigcup_{\delta=0}^{d} \Delta^\delta$ $(\Delta^0 = \{\vec{\emptyset}\})$.
- The $(\vec{i}, \vec{j})$-element of C is defined as follows.

$$C_{\vec{i},\vec{j}} = \begin{cases} c_{\vec{i},\vec{j}}, & \text{if } |\vec{i}| = |\vec{j}|\text{, that is, } (\vec{i}, \vec{j}) \in \Delta^\delta \times \Delta^\delta \text{ for some } \delta; \\ 0, & \text{otherwise.} \end{cases}$$

In the remainder of this paper, we refer to C as a *coefficient matrix* of p.

When $\deg p \le 1$ or $D = 1$, a coefficient matrix is unique for each polynomial p. In contrast, when $\deg p \ge 2$ and $D \ge 2$, there may exist more than one coefficient matrices for the same p.

Example 3. If $\deg p = 2$ and $D = 2$, p can include 15 mutually independent coefficients, that is, $a, b_1, \dots, b_4, c_1, \dots, c_9$ and c_{10} as shown by (11).

$$\begin{aligned} p[\xi_{11}, \xi_{12}, \xi_{21}, \xi_{22}] &= a + b_1\xi_{11} + b_2\xi_{12} + b_3\xi_{21} + b_4\xi_{22} + c_1\xi_{11}^2 + c_2\xi_{11}\xi_{12} + c_3\xi_{12}^2 \\ &+ c_4\xi_{11}\xi_{21} + c_5\xi_{11}\xi_{22} + +c_6\xi_{12}\xi_{21} + c_7\xi_{12}\xi_{22} + c_8\xi_{21}^2 + c_9\xi_{21}\xi_{22} + c_{10}\xi_{22}^2. \end{aligned} \tag{11}$$

On the other hand, a coefficient matrix of p essentially includes 21 $(= 1 + 2^2 + 4^2)$ variable elements as shown below.

$$p = c_{\vec{\emptyset},\vec{\emptyset}} + \sum_{i\in\{0,1\}}\sum_{j\in\{0,1\}} c_{(i),(j)}\xi_{ij} + \sum_{(i_1,i_2)\in\{0,1\}^2}\sum_{(j_1,j_2)\in\{0,1\}^2} c_{(i_1,i_2),(j_1,j_2)}\xi_{i_1j_1}\xi_{i_2j_2}. \tag{12}$$

This implies that, for the fixed p, the set of the coefficient matrices that represent p is of dimension 6 $(= 21 - 15)$ in the entire 21-dimensional linear space of coefficient matrices.

We will look at this more closely. In comparing the formulas (11) and (12) in the terms of degree smaller than or equal to 1, we have the following deterministic equations in $c_{\vec{i},\vec{j}}$.

$$c_{\vec{\emptyset},\vec{\emptyset}} = a, \quad c_{(1),(1)} = b_1, \quad c_{(1),(2)} = b_2, \quad c_{(2),(1)} = b_3, \quad c_{(2),(2)} = b_4.$$

In contrast, the same comparison with respect to the terms of degree 2 yields only a set of indeterministic equations in $c_{\vec{i},\vec{j}}$. For example, the comparison in the term $\xi_{11}\xi_{12}$, that is, $c_{(1,1),(1,2)}\xi_{11}\xi_{12} + c_{(1,1),(2,1)}\xi_{12}\xi_{11} = c_2\xi_{11}\xi_{12}$ yields $c_{(1,1),(1,2)} + c_{(1,1),(2,1)} = c_2$.

Finally, we see that the set of the matrices $C_{\alpha,\beta,\gamma,\delta,\epsilon,\varphi}$ including 6 real parameters α, β, γ, δ, ϵ and φ is exactly the set of the coefficient matrices of p.

$$C_{\alpha,\beta,\gamma,\delta,\epsilon,\varphi} = \begin{bmatrix} a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_1 & b_2 & 0 & 0 & 0 & 0 \\ 0 & b_3 & b_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_1 & \alpha & c_2 - \alpha & c_3 \\ 0 & 0 & 0 & \beta & \gamma & \delta & \epsilon \\ 0 & 0 & 0 & c_4 - \beta & c_6 - \delta & c_5 - \gamma & c_7 - \epsilon \\ 0 & 0 & 0 & c_8 & \varphi & c_9 - \varphi & c_{10} \end{bmatrix}.$$

5.2. Statement of the main theorem

The main theorem of this paper is presented as follows.

Theorem 9. *Let p be a real polynomial in the D^2 variables of $\{\xi_{ij} \mid (i,j) \in \{1, \dots, D\}^2\}$. If there exists a positive semidefinite coefficient matrix for p, the p-summary $p[k]$ of an arbitrary positive semidefinite underlying kernel k is positive semidefinite.*

All the polynomials don't have a positive semidefinite coefficient matrix.

Example 4. In Example 3, if one of $a, b_1, b_4, c_1, c_3, c_5$ and c_{10} is negative, the coefficient matrix $C_{\alpha,\beta,\gamma,\delta,\epsilon,\varphi}$ has at least one negative diagonal element. Hence, such p cannot have any positive semidefinite coefficient matrix.

Also, a polynomial p may have an infinite number of positive semidefinite coefficient matrices. This is because the condition for the existence of positive semidefinite coefficient matrices is given by a set of algebraic inequalities in the coefficients of p.

5.3. Key lemma

Before we sketch a plot of the proof of Theorem 9 given in the present paper, we introduce Lemma 11, which will play a key role in the proof. A proof of Lemma 11 is given in Section 6.1.

Let X^{ij} be m-dimensional square matrices parameterized by $(i,j) = \{1, \dots, n\}^2$, and let X denote the derived mn-dimensional square matrix $[X^{ij}]_{i,j=1,\dots,n}$: the $(m(i-1)+k, m(j-1)+l)$-element of X, denoted by X^{ij}_{kl}, is defined to be the (k,l)-element of X^{ij}.

Definition 10. For an m-dimensional square matrix A, the n-dimensional square matrix $\left[\mathrm{tr}({}^{\top}AX^{ij})\right]_{i,j=1,\dots,n} = \left[\sum_{k=1}^m\sum_{l=1}^m A_{kl}X^{ij}_{kl}\right]_{i,j=1,\dots,n}$ is called the *A-linear summary matrix* of X, and is denoted by $\mathrm{smry}_A(X)$.

Lemma 11. *For an m-dimensional real matrix A, the following are equivalent to each other.*

(1) *A is positive semidefinite.*
(2) *The linear summary matrix $\mathrm{smry}_A(X)$ is positive semidefinite for an arbitrary mn-dimensional positive semidefinite matrix X.*

5.4. A plot of the proof of Theorem 9

Now, we are ready to sketch a plot of the proof of Theorem 9 that is given in 6.2. The objective of the proof is to show that arbitrary Gram matrices G of $p[k]$ are positive semidefinite (Definition 3), if p has a positive semidefinite coefficient matrix.

In the remainder of this subsection, we assume the following restriction.

- $D = 2$;
- The polynomial $p(\xi_{11}, \xi_{12}, \xi_{21}, \xi_{22})$ is homogeneous of degree 1 or 2.
- The underlying kernel k is positive semidefinite.
- The Gram matrix G is for two data points $x^{(1)} = (x_1^{(1)}, x_2^{(1)})$ and $x^{(2)} = (x_1^{(2)}, x_2^{(2)})$; Therefore, G is given as follows;

$$G = \begin{bmatrix} p[k](x^{(1)}, x^{(1)}), & p[k](x^{(1)}, x^{(2)}) \\ p[k](x^{(2)}, x^{(1)}), & p[k](x^{(2)}, x^{(2)}) \end{bmatrix}.$$

5.4.1. Case $\deg(p) = 1$:

Let p, C and X be as follows. Since p is assumed to be homogeneous, we omit the row and columns corresponding to the constant term $c_{\vec{\emptyset},\vec{\emptyset}}$ of p. Therefore, C is of dimension 4 instead of 5.

$$p(\xi_{11}, \xi_{12}, \xi_{21}, \xi_{22}) = c_{11}\xi_{11} + c_{12}\xi_{12} + c_{21}\xi_{21} + c_{22}\xi_{22}$$

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}, \quad X^{ij} = \begin{bmatrix} k(x_1^{(i)}, x_1^{(j)}) & k(x_1^{(i)}, x_2^{(j)}) \\ k(x_2^{(i)}, x_1^{(j)}) & k(x_2^{(i)}, x_2^{(j)}) \end{bmatrix}, \quad X = \begin{bmatrix} X^{11} & X^{12} \\ X^{21} & X^{22} \end{bmatrix}.$$

The important relation to be noted here is the following.

$$\begin{aligned} p[k](x^{(i)}, x^{(j)}) &= c_{11}k(x_1^{(i)}, x_1^{(j)}) + c_{12}k(x_1^{(i)}, x_2^{(j)}) + c_{21}k(x_2^{(i)}, x_1^{(j)}) \\ &\quad + c_{22}k(x_2^{(i)}, x_2^{(j)}) = C^{\mathsf{T}} X^{ij} \end{aligned}$$

$$\therefore \quad G = \mathrm{smry}_C(X).$$

The 4-dimensional matrix X is positive semidefinite, since it is a Gram matrix with respect to the underlying kernel k. Hence, the claim that G is positive semidefinite immediately follows from Lemma 11.

5.4.2. Case $\deg(p) = 2$:

Let p, C, X and Y be as follows.

$$p(\xi_{11}, \xi_{12}, \xi_{21}, \xi_{22}) = \sum_{k_1=1}^{2}\sum_{k_2=1}^{2}\sum_{l_1=1}^{2}\sum_{l_2=1}^{2} c_{(k_1,k_2),(l_1,l_2)}\xi_{k_1 l_1}\xi_{k_2 l_2}$$

$$C_{2(k_1-1)+k_2, 2(l_1-1)+l_2} = c_{(k_1,k_2),(l_1,l_2)}, \quad C = \left[C_{2(k_1-1)+k_2,2(l_1-1)+l_2}\right]_{k_1,k_2,l_1,l_2=1,2}$$

$$X^{ij} = \begin{bmatrix} k(x_1^{(i)}, x_1^{(j)}) & k(x_1^{(i)}, x_2^{(j)}) \\ k(x_2^{(i)}, x_1^{(j)}) & k(x_2^{(i)}, x_2^{(j)}) \end{bmatrix}, \quad X = \begin{bmatrix} X^{11} & X^{12} \\ X^{21} & X^{22} \end{bmatrix}$$

$$Y^{ij}_{2(k_1-1)+k_2, 2(l_1-1)+l_2} = k(x_{k_1}^{(i)}, x_{l_1}^{(j)})k(x_{k_2}^{(i)}, x_{l_2}^{(j)})$$

$$Y^{ij} = \left[Y^{ij}_{2(k_1-1)+k_2,2(l_1-1)+l_2}\right]_{k_1,k_2,l_1,l_2=1,2}, \quad Y = \begin{bmatrix} Y^{11} & Y^{12} \\ Y^{21} & Y^{22} \end{bmatrix}.$$

In the same way as seen in 5.4.1, $G = \mathrm{smry}_C(Y)$ holds. However, to apply Lemma 11 and to conclude that G is positive semidefinite, we need to claim that Y is positive semidefinite.

The claim is proved as follows. When we define the tensor product matrix $X \otimes X$ as follows, we have $Y = (X \otimes X)[\{10a + 4b + c + 1 \mid a, b, c = 0, 1\}]$. Note that $\xi \otimes \xi$ is of dimension 16.

$$X \otimes X = \begin{bmatrix} k(x_1^{(1)}, x_1^{(1)})X, & k(x_1^{(1)}, x_2^{(1)})X, & k(x_1^{(1)}, x_1^{(2)})X, & k(x_1^{(1)}, x_2^{(2)})X \\ k(x_2^{(1)}, x_1^{(1)})X, & k(x_2^{(1)}, x_2^{(1)})X, & k(x_2^{(1)}, x_1^{(2)})X, & k(x_2^{(1)}, x_2^{(2)})X \\ k(x_1^{(2)}, x_1^{(1)})X, & k(x_1^{(2)}, x_2^{(1)})X, & k(x_1^{(2)}, x_1^{(2)})X, & k(x_1^{(2)}, x_2^{(2)})X \\ k(x_2^{(2)}, x_1^{(1)})X, & k(x_2^{(2)}, x_2^{(1)})X, & k(x_2^{(2)}, x_1^{(2)})X, & k(x_2^{(2)}, x_2^{(2)})X \end{bmatrix}$$

$X \otimes X$ is positive semidefinite, since so is X (*e.g.* [2]). Therefore, Y is positive semidefinite by Proposition 5.

6. A proof of Theorem 9

In this section, we will prove Lemma 11 (6.1) and Theorem 9 (6.2).

6.1. A proof of Lemma 11

We first prove Lemma 12.

Lemma 12. *For an m-dimensional real matrix A, the following are equivalent to each other.*

(1) *A is symmetric.*
(2) *The linear summary matrix* $\mathrm{smry}_A(X)$ *is symmetric for an arbitrary mn-dimensional symmetric matrix X.*
(3) *The linear summary matrix* $\mathrm{smry}_A(X)$ *is symmetric for an arbitrary mn-dimensional positive semidefinite matrix X.*

Proof. (1) ⇒(2): The assertion follows from $X^{ji} = X^{ij\mathsf{T}}$ and $\mathrm{tr}(YZ) = \mathrm{tr}(Y^{\mathsf{T}}Z^{\mathsf{T}})$.

$$\mathrm{tr}({}^{\mathsf{T}}AX^{ij}) = \mathrm{tr}(AX^{ij\mathsf{T}}) = \mathrm{tr}(AX^{ji}) = \mathrm{tr}({}^{\mathsf{T}}AX^{ji})$$

(2) ⇒(3): The assertion is apparent, since a positive semidefinite matrix is symmetric by definition.
(3) ⇒(1): For arbitrary $\alpha, \beta \in \{1, \ldots, m\}$, define $X = [X^{ij}]_{i,j=1,2}$ as follows.

$$X^{ij}_{kl} = \begin{cases} 1, & \text{if } (i, j, k, l) = (1, 1, \alpha, \alpha), (1, 2, \alpha, \beta) \\ & \qquad\qquad (2, 1, \beta, \alpha), (2, 2, \beta, \beta) \\ 0, & \text{otherwise.} \end{cases}$$

Then, $\mathrm{smry}_A(X)$ is as follows.

$$\mathrm{smry}_A(X) = \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha\beta} \\ A_{\beta\alpha} & A_{\beta\beta} \end{bmatrix}.$$

Since X is positive semidefinite, $\mathrm{smry}_A(X)$ is symmetric. Hence, $A_{\alpha\beta} = A_{\beta\alpha}$ holds for arbitrary α and β. □

To prove the remainder of the assertion of Lemma 11, we first prove it for the case where A is diagonal, and then prove it for the general case.

Assume that A is a positive semidefinite diagonal matrix.

$$A = \begin{bmatrix} \alpha_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \alpha_m \end{bmatrix}, \quad \alpha_I \geq 0 \quad \text{for } I = 1, \ldots, m.$$

If X is positive semidefinite, there exists an mn-dimensional square matrix $Y = [Y^{ij}]_{i,j=1,\ldots,n}$ such that $X = Y^{\mathsf{T}}Y$ by Proposition 1.

$$\mathrm{tr}\,{}^{\mathsf{T}}AX^{ij} = \sum_{I=1}^{m} \alpha_I \left(\sum_{k=1}^{n} \sum_{l=1}^{m} Y^{ki}_{lI} Y^{kj}_{lI} \right) = \sum_{I=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{m} (\sqrt{\alpha_I} Y^{ki}_{lI})(\sqrt{\alpha_I} Y^{kj}_{lI}).$$

Therefore, $\mathrm{smry}_A(X) = Z^{\mathsf{T}}Z$ holds for the $m^2 n \times n$ matrix Z such that

$$Z_{mn(I-1)+m(k-1)+l,i} = \sqrt{\alpha_I} Y^{ki}_{lI}.$$

Therefore, $\mathrm{smry}_A(X)$ is positive semidefinite by Proposition 1.

To prove the inverse, we assume $n = 1$. For $I \in \{1, \ldots, m\}$, we let $X^{(I)}$ be the m-dimensional positive semidefinite matrix whose elements are 0 except for $X^{(I)}_{II} = 1$. Since $\mathrm{smry}_A(X^{(I)}) = \alpha_I \geq 0$ holds, A turns out positive semidefinite.

Now, we claim that the assertion for the general case where A is not necessarily diagonal is reduced to that of the diagonal case. Note that A is symmetric by Lemma 12. Therefore, $P^{\mathsf{T}}AP$ is diagonal for some orthogonal matrix P. Our claim immediately follows from the properties shown below.

(1) A is positive semidefinite, if, and only if, so is $P^{\mathsf{T}}AP$.
(2) X is positive semidefinite, if, and only if, so is $Y = \left[P^{\mathsf{T}}X^{ij}P\right]_{i,j=1,\ldots n}$.
(3) $\mathrm{smry}_A(X) = \mathrm{smry}_{P^{\mathsf{T}}AP}(Y)$, since

$$\mathrm{tr}({}^{\mathsf{T}}AX^{ij}) = \mathrm{tr}\left(P^{\mathsf{T}}({}^{\mathsf{T}}AX^{ij})P\right) = \mathrm{tr}\left((P^{\mathsf{T}}AP)^{\mathsf{T}}(P^{\mathsf{T}}X^{ij}P)\right).$$

Now, we have completed the proof of Lemma 11.

As a corollary to Lemma 11, we have the following characterization of the positive semidefinite matrix.

Corollary 13. *For an m-dimensional real matrix A, the following are equivalent to each other.*

(1) *A is positive semidefinite.*
(2) *A is a symmetric matrix such that* $\mathrm{tr}({}^{\mathsf{T}}AX) \geq 0$ *for an arbitrary m-dimensional positive semidefinite matrix X.*

Proof. The assertion that (1) implies (2) immediately follows from Lemma 11.

Conversely, we assume that (2) holds. Since A is symmetric, $P^{\mathsf{T}}AP$ is diagonal for some orthogonal P. When α_I denotes the I-th diagonal element of $P^{\mathsf{T}}AP$, the assertion follows from the fact that $\alpha_I = \mathrm{tr}({}^{\mathsf{T}}AX^{(I)}) \geq 0$. □

6.2. A proof of Theorem 9

The notion of the linear summaries of matrices can be naturally extended to that of the *polynomial summaries* of matrices, and we have Lemma 15 as a generalization of Lemma 11.

Definition 14. Let p be a polynomial of degree d in the m^2 variables $\xi_{11}, \xi_{12}, \ldots, \xi_{mm}$. The *p-polynomial summary matrix* $\mathrm{smry}_p[X]$ of an mn-dimensional matrix $X = [X^{ij}]_{i,j=1,\ldots,n}$ is the n-dimensional matrix defined as follows.

$$\mathrm{smry}_p[X] = \left[p(X^{ij}_{11}, X^{ij}_{12}, \ldots, X^{ij}_{mm})\right]_{i,j=1,\ldots,n}.$$

If p is linear with the unique coefficient matrix C, $\mathrm{smry}_p[X]$ is identical to $\mathrm{smry}_C(X)$.

Lemma 15. *Let p be a real polynomial in the m^2 variables $\{\xi_{ij}\}_{i,j=1,\ldots,m}$. If p has a positive semidefinite coefficient matrix C, $\mathrm{smry}_p[X]$ is positive semidefinite for an arbitrary mn-dimensional positive semidefinite matrix X.*

Proof. The symmetry property of $\mathrm{smry}_p[X]$ immediately follows from those of C and X.

For $\delta \in \{0, \ldots, d\}$ and $(i, j) \in \{1, \ldots, n\}^2$, we define the m^δ-dimensional matrix $\bar{X}\langle\delta\rangle^{ij}$ as follows.

- The rows and columns are indexed by the vector $\vec{k} \in \{1, \ldots, m\}^\delta$.
- For $\vec{k} = (k_1, \ldots, k_\delta)$ and $\vec{l} = (l_1, \ldots, l_\delta)$, the $(\vec{k}, \vec{l})$-element is:

$$\bar{X}\langle\delta\rangle^{ij}_{\vec{k},\vec{l}} = \prod_{\alpha=1}^{\delta} X^{ij}_{k_\alpha l_\alpha}.$$

In particular, we let $\bar{X}\langle 0\rangle^{ij} = [1]$ for $\delta = 0$.

Then the $m^\delta n$-dimensional square matrix $\bar{X}\langle\delta\rangle$ is defined using $\bar{X}\langle\delta\rangle^{ij}$.

$$\bar{X}\langle\delta\rangle = \begin{bmatrix} \bar{X}\langle\delta\rangle^{11} & \ldots & \bar{X}\langle\delta\rangle^{1n} \\ \vdots & \ddots & \vdots \\ \bar{X}\langle\delta\rangle^{n1} & \ldots & \bar{X}\langle\delta\rangle^{nn} \end{bmatrix}.$$

On the other hand, we can define the matrix $X^{\otimes\delta}$ for the δ-th power tensor product of X as follows.

- The rows and columns are indexed by the pair of vectors $(\vec{i}, \vec{k}) \in \{1, \ldots, n\}^\delta \times \{1, \ldots, m\}^\delta$.
- For $(\vec{i}, \vec{k}) = ((i_1, \ldots, i_\delta), (k_1, \ldots, k_\delta))$ and $(\vec{j}, \vec{l}) = ((j_1, \ldots, j_\delta), (l_1, \ldots, l_\delta))$, the $((\vec{i}, \vec{k}), (\vec{j}, \vec{l}))$-element $\left(\bar{X}^{\otimes\delta}\right)_{(\vec{i},\vec{k}),(\vec{j},\vec{l})}$ is:

$$\left(\bar{X}^{\otimes\delta}\right)_{(\vec{i},\vec{k}),(\vec{j},\vec{l})} = \prod_{\alpha=1}^{\delta} X^{i_\alpha j_\alpha}_{k_\alpha l_\alpha}.$$

Since $X^{\otimes\delta}$ is positive semidefinite (*e.g.* [1]) and

$$X\langle\delta\rangle = X^{\otimes\delta}[\{(\underbrace{i, \ldots, i}_{\delta}) \mid i = 1, \ldots, n\} \times \{1, \ldots, m\}^\delta]$$

holds, we see that $\bar{X}\langle\delta\rangle$ proves positive semidefinite by Proposition 5.

Hence, the assertion follows from

$$\mathrm{smry}_p[X] = \mathrm{smry}_C(\bigoplus_{\delta=0}^{d} \bar{X}\langle\delta\rangle)$$

by Lemma 11. □

Apparently, Theorem 9 is a direct corollary to Lemma 15.

7. The converse of Theorem 9 for the polynomials of degree 1

In this section, we show that the reverse of Theorem 9 holds true for an arbitrary polynomial p of degree 1. Moreover, we will see that, if the unique coefficient matrix of p is not positive semidefinite, $p[k]$ is not positive semidefinite for some positive semidefinite underlying kernel $k : \chi_* \times \chi_*$ defined over a small space χ_*.

Let a polynomial $p(\xi_{11}, \xi_{12}, \ldots, \xi_{DD})$ be of degree 1.

$$p = c_{\vec{\emptyset},\vec{\emptyset}} + \sum_{i=1}^{D} \sum_{j=1}^{D} c_{(i),(j)} \xi_{ij}.$$

If the constant term $c_{\vec{\emptyset},\vec{\emptyset}}$ is negative, $p[\epsilon k]((x_1, \ldots, x_D), (x_1, \ldots, x_D)) < 0$ holds for an arbitrary positive semidefinite underlying kernel $k(x_i, x_j)$ and a sufficiently small $\epsilon > 0$. Hence, we have $c_{\vec{\emptyset},\vec{\emptyset}} \geq 0$.

Also, we claim that the submatrix $C_1 = \left[c_{(i),(j)}\right]_{i,j=1,\ldots,D}$ of the coefficient matrix of p is positive semidefinite. We will prove the contraposition of the claim. If C_1 is not positive semidefinite, Corollary 13 asserts that there exists a D-dimensional positive semidefinite matrix X such that $\mathrm{tr}({C_1}^{\mathsf{T}} X) < 0$. We define a positive semidefinite kernel k over $\chi_* = \{x_1, \ldots, x_D\}$ by $k(x_i, x_j) = \gamma X_{ij}$. Therefore, for a sufficiently large positive γ, we have the following.

$$p[k]((x_1, \ldots, x_D), (x_1, \ldots, x_D)) = c_{\emptyset,\emptyset} + \gamma C^{\mathsf{T}}{}_1 X < 0.$$

Thus, $p[k]$ is not positive semidefinite, and we finally obtain the following proposition.

Proposition 16. *Assume that $p(\xi_{11}, \xi_{12}, \ldots, \xi_{DD})$ is a polynomial of degree 1 such that its unique coefficient matrix is not positive semidefinite. Then, there exists χ_* and $k : \chi_* \times \chi_* \longrightarrow \mathbb{R}$ such that $|\ \chi_*\ | \leq D$, k is positive semidefinite, and $p[k]$ is not positive semidefinite.*

8. Applications

8.1. Generalization of polynomial kernels

The following direct corollary to Theorem 9 presents the multivariate version of the polynomial kernels (3.1.1).

Corollary 17. *Let p be a real polynomial in the D variables of $\{\xi_{ii} \mid i = 1, \ldots, D\}$. If p includes only non-negative coefficients, then the p-summary $p[k]$ of a positive semidefinite underlying kernel k is always positive semidefinite.*

Proof. The polynomial p has a diagonal coefficient matrix with non-negative diagonal elements, which is apparently positive semidefinite. Hence, we have the assertion by Theorem 9. □

Apparently, the polynomial kernel is the special case for $D = 1$ of the polynomial summary of Corollary 17.

8.2. Generalization of the determinant kernels

Let $x = (x_1, \ldots, x_D)$ and $y = (y_1, \ldots, y_D)$ be two data points in χ_*^D and $\Phi(\xi)$ be the characteristic polynomial of the Gram matrix G with respect to $k : \chi_* \times \chi_* \to \mathbb{R}$.

$$G = \begin{bmatrix} k(x_1, y_1) & \ldots & k(x_1, y_D) \\ \vdots & \ddots & \vdots \\ k(x_D, y_1) & \ldots & k(x_D, y_D) \end{bmatrix}.$$

Wolf et al. [7] and Zhou [9] proved that the constant term of $\Phi(\xi)$, that is $\det(G)$, determines a positive semidefinite kernel, if k is positive semidefinite.

Generalizing the result, we will see that all of the coefficients of $\Phi(\xi)$ determine positive semidefinite kernels. When $(-1)^d \Phi_{D,d}$ denotes the coefficient of the term of ξ^d of Eq. (13), $\Phi_{D,d}$ is a $(D - d)$-degree homogeneous polynomial in $\{\xi_{11}, \xi_{12}, \ldots, \xi_{DD}\}$.

$$\det \begin{bmatrix} \xi_{11} - \xi & \ldots & \xi_{1D} \\ \vdots & \ddots & \vdots \\ \xi_{D1} & \ldots & \xi_{DD} - \xi \end{bmatrix}. \tag{13}$$

Thus, we have the following.

$$\Phi(\xi) = \sum_{d=0}^{D} (-1)^d \left(\Phi_{D,d}[k](x, y)\right) \xi^d.$$

In particular, the following holds.

$$\mathrm{tr}(G) = \Phi_{D,D-1}[k](x, y)$$
$$\det(G) = \Phi_{D,0}[k](x, y).$$

Thus, Corollary 18 is a generalization of the results of [7] and [9].

Corollary 18. *For an arbitrary positive semidefinite underlying kernel k, $\Phi_{D,d}[k](x,y)$ are positive semidefinite kernels for $d = 0, \ldots, D$.*

Proof. $\Phi_{D,d}(\xi_{11}, \xi_{12}, \ldots, \xi_{DD})$ is evaluated as follows, where $\mathfrak{S}_{D-d}$ denotes the permutation group acting on $\{1, \ldots, D-d\}$ and $\mathrm{sgn}(\pi)$ does the sign of the permutation $\pi \in \mathfrak{S}_{D-d}$.

$$\begin{aligned}\Phi_{D,d} &= \sum_{1\le\alpha_1<\cdots<\alpha_{D-d}\le D} \det([\xi_{\alpha_i\alpha_j}]_{i,j=1,\ldots,D-d}) \\ &= \sum_{1\le\alpha_1<\cdots<\alpha_{D-d}\le D} \sum_{\pi\in\mathfrak{S}_{D-d}} \mathrm{sgn}(\pi) \prod_{i=1}^{D-d} \xi_{\alpha_i\alpha_{\pi(i)}}.\end{aligned}$$

We fix an instance of $1 \le \alpha_1 < \cdots < \alpha_{D-d} \le D$, and show that the coefficient matrix $C_{\alpha_1,\ldots,\alpha_{D-d}}$ for $\sum_{\pi\in\mathfrak{S}_{D-d}} \mathrm{sgn}(\pi) \prod_{i=1}^{D-d} \xi_{\alpha_i\alpha_{\pi(i)}}$ is positive semidefinite.

$$\begin{aligned}\sum_{\pi\in\mathfrak{S}_{D-d}} \mathrm{sgn}(\pi) \prod_{i=1}^{D-d} \xi_{\alpha_i\alpha_{\pi(i)}} &= \sum_{\pi\in\mathfrak{S}_{D-d}} \sum_{\sigma\in\mathfrak{S}_{D-d}} \frac{\mathrm{sgn}(\sigma)\mathrm{sgn}(\pi\circ\sigma)}{(D-d)!} \prod_{i=1}^{D-d} \xi_{\alpha_{\sigma(i)}\alpha_{\pi(\sigma(i))}} \\ &= \sum_{\sigma\in\mathfrak{S}_{D-d}} \sum_{\tau\in\mathfrak{S}_{D-d}} \frac{\mathrm{sgn}(\sigma)}{\sqrt{(D-d)!}} \frac{\mathrm{sgn}(\tau)}{\sqrt{(D-d)!}} \prod_{i=1}^{D-d} \xi_{\alpha_{\sigma(i)}\alpha_{\tau(i)}}.\end{aligned}$$

Therefore, $C_{\alpha_1,\ldots,\alpha_{D-d}}$ is equal to $\vec{c}^{\top}\vec{c}$ for the row vector $\vec{c} = \left(\frac{\mathrm{sgn}(\sigma)}{\sqrt{(D-d)!}}\right)_{\sigma\in\mathfrak{S}_{D-d}}$, and therefore, is positive semidefinite. □

8.3. Correction to the CI and WDwS kernels

In this subsection, just for simplicity, we assume $n = 1$ in Eq. (6). We have a generic technique to generalize the result for $n = 1$ to the case of $n > 1$ [5].

First, we fix $s \in \{1, \ldots, S\}$, and determine a sufficient condition for $K(x,y)$ of Eq. (14) to be positive semidefinite.

$$K(x,y) = \sum_{i=1}^{L} w_i \left[k(x_i,y_i) + \bar{w}_s \{k(x_{i+s},y_i) + k(x_i,y_{i+s})\}\right]. \tag{14}$$

For non-negative integers a and b such that $a \in \{1, \ldots, s\}$ and $s(b-1)+a \le L$, we define $\gamma_b^{(a)}$ by the recurrence formulas described below.

$$\gamma_0^{(a)} = 1, \quad \gamma_1^{(a)} = w_a, \quad \gamma_b^{(a)} = w_{s(b-1)+a}\gamma_{b-1}^{(a)} - \bar{w}_s^2 w_{s(b-2)+a}^2 \gamma_{b-2}^{(a)}. \tag{15}$$

Then, we obtain the following result as a corollary to Theorem 9 and Proposition 16.

Corollary 19. *If $\gamma_b^{(a)} > 0$ holds for every (a,b) such that $a \in \{1, \ldots, s\}$ and $s(b-1)+a \le L$, $K(x,y)$ defined by Eq. (14) is positive semidefinite for an arbitrary positive semidefinite underlying kernel $k(x_i,y_j)$.*

Conversely, if $\gamma_b^{(a)} < 0$ holds for some (a,b), there exists a positive semidefinite kernel $k(x_i,y_j)$ such that the resulting $K(x,y)$ is not positive semidefinite.

Proof. When we define $c_{i,j}$ by: $c_{i,j} = w_i$, if $i = j$; $c_{i,j} = \bar{w}_s w_i$, if $j = i+s$; $c_{i,j} = \bar{w}_s w_j$, if $i = j+s$; and $c_{i,j} = 0$, otherwise. Then, $K(x,y) = \sum_{i=1}^{L}\sum_{j=1}^{L} c_{i,j} k(x_i,y_j)$ holds for $K(x,y)$ of Eq. (14). Therefore, by Theorem 9, it suffices to show that the matrix $C = [c_{i,j}]_{(i,j)\in\{1,\ldots,L\}^2}$ is positive semidefinite to prove the first assertion of Corollary 19.

We let a be one of $\{1, 2, \ldots, s\}$, and let the submatrix $C_b^{(a)}$ denote the b-dimensional matrix $[c_{s(i-1)+a,s(j-1)+a}]_{(i,j)=\{1,\ldots,b\}^2}$. For $L = sq + r$ such that $r \in \{1, 2, \ldots, s\}$, b moves in the interval $[1, q+1]$ if $a \le r$, and does in the interval $[1, q]$ if $a > r$. For example, when $b = 4$, $C_4^{(a)}$ looks as follows.

$$C_4^{(a)} = \begin{bmatrix} w_a & \bar{w}_s w_a & 0 & 0 \\ \bar{w}_s w_a & w_{s+a} & \bar{w}_s w_{s+a} & 0 \\ 0 & \bar{w}_s w_{s+a} & w_{2s+a} & \bar{w}_s w_{2s+a} \\ 0 & 0 & \bar{w}_s w_{2s+a} & w_{3s+a} \end{bmatrix}.$$

By applying the same permutation to the rows and the columns of C if necessary, C is decomposed into a direct sum of its sub-matrices as follows.

$$C = C_{q+1}^{(1)} \oplus \cdots \oplus C_{q+1}^{(r)} \oplus C_q^{(r+1)} \oplus \cdots \oplus C_q^{(s)}.$$

Therefore, C is positive definite (*i.e.* C has only positive eigenvalues), if, and only if, so are $C_{q+1}^{(a)}$ for $a \le r$ and $C_q^{(a)}$ for $a > r$.

On the other hand, by Proposition 4, $C_{q+1}^{(a)}$ for $a \leq r$ (resp. $C_q^{(a)}$ for $a > r$) is positive definite, if, and only if, $\det(C_b^{(a)}) > 0$ for all $1 \leq b \leq q+1$ (resp. $1 \leq b \leq q$). By the Laplacian determinant expansion by minors, we have the recurrence formula (16). This indicates that $\det(C_b^{(a)})$ coincides with $\gamma_b^{(a)}$. Thus, the first assertion of Corollary 19 has been proved.

$$\det(C_0^{(a)}) = 1, \quad \det(C_1^{(a)}) = w_a$$

$$\det(C_b^{(a)}) = w_{(b-1)s+a} \det(C_{b-1}^{(a)}) - (w_{(b-2)s+a}\bar{w}_s)^2 \det(C_{b-2}^{(a)}). \tag{16}$$

The second assertion is also derived from Corollary 16. If $\gamma_b^{(a)} < 0$ for some (a, b), $\det(C_b^{(a)})$ is negative, and hence $C_b^{(a)}$ is not positive semidefinite. Hence, there exists a positive semidefinite kernel $k(x_i, y_j)$ defined over the alphabet Σ such that $K(x, y)$ is not positive semidefinite. □

The sufficient condition presented in Corollary 19 is *very close* to a necessary condition, since the positive semidefiniteness of $K(x, y)$ is left undetermined only in the marginal cases where $\gamma_b^{(a)} \geq 0$ for all (a, b) and $\gamma_b^{(a)} = 0$ for some (a, b).

On the other hand, when $w_1, \ldots, w_L$ are fixed, the condition is reduced to an equivalent inequality of $0 \leq \bar{w}_s < b^{(s)}_{w_1,\ldots,w_L}$ for some $b^{(s)}_{w_1,\ldots,w_L}$. While it is not easy to determine the actual values for $b^{(s)}_{w_1,\ldots,w_L}$, Corollary 20 gives an easily computable lower bound for $b^{(s)}_{w_1,\ldots,w_L}$.

Corollary 20. *Assume that all the weights are positive. The kernel K defined by Eq. (14) is positive semidefinite for an arbitrary positive semidefinite $k(x_i, y_j)$, if the following inequality holds for $w_1, \ldots, w_L$ and $\bar{w}_s$.*

$$\bar{w}_s \leq \min\left\{ \frac{w_i}{w_{i-s} + w_i} \middle| i = s+1, \ldots, L \right\}. \tag{17}$$

Proof. Assuming that $\bar{w}_s$ satisfies the inequality (17), we will prove that $\gamma_b^{(a)} > 0$ holds for an arbitrary pair of non-negative integers (a, b) such that $a \in \{1, \ldots, s\}$ and $s(b-1) + a \leq L$.

To start with, we define $B_b^{(a)}$ as the matrix obtained by replacing the (b, b)-element $w_{s(b-1)+a}$ of $C_b^{(a)}$ with $\bar{w}_s w_{s(b-1)+a}$, and let $\beta_b^{(a)} = \det(B_b^{(a)})$. For example, $B_4^{(a)}$ looks as follows (compare with $C_4^{(a)}$).

$$B_4^{(a)} = \begin{bmatrix} w_a & \bar{w}_s w_a & 0 & 0 \\ \bar{w}_s w_a & w_{s+a} & \bar{w}_s w_{s+a} & 0 \\ 0 & \bar{w}_s w_{s+a} & w_{2s+a} & \bar{w}_s w_{2s+a} \\ 0 & 0 & \bar{w}_s w_{2s+a} & \bar{w}_s w_{3s+a} \end{bmatrix}.$$

In the rest of this section, we fix $a \in \{1, \ldots, s\}$, and prove $\gamma_b^{(a)} > 0$ and $\beta_b^{(a)} > 0$ by induction on b. Furthermore, we can assume $b > 1$, since $\gamma_1^{(a)} = w_a > 0$ and $\beta_1^{(a)} = \bar{w}_s w_a > 0$ hold.

First, we confirm a few key properties.

- The hypothesis (17) implies $\bar{w}_s < 1$.
- Therefore, $\gamma_b^{(a)} > \beta_b^{(a)}$ follows from the hypothesis of induction $\gamma_{b-1}^{(a)} > 0$. This implies that we have only to show $\beta_b^{(a)} > 0$ to complete the proof.
- The inequality $\bar{w}_s \leq w_{s(b-1)+a}/(w_{s(b-2)+a} + w_{s(b-1)+a})$ implies the following.

$$1 - \frac{\bar{w}_s w_{s(b-2)+a}}{w_{s(b-1)+a}} \geq 1 - \bar{w}_s \left(\frac{1}{\bar{w}_s} - 1 \right) = \bar{w}_s. \tag{18}$$

To show $\beta_b^{(a)} > 0$, we first expand $\beta_b^{(a)}$ and $\gamma_{b-1}^{(a)}$ by Laplacian determinant expansion, apply the inequality of (18) (note that $\gamma_{b-2}^{(a)} > 0$ holds by the hypothesis of induction), and then collect up the terms into $\beta_{b-1}^{(a)}$ by applying Laplacian determinant expansion in reverse. The assertion follows from the hypothesis of induction $\beta_{b-1}^{(a)} > 0$.

$$\begin{aligned} \beta_b^{(a)} &= \bar{w}_s w_{s(b-1)+a} \gamma_{b-1}^{(a)} - (\bar{w}_s w_{s(b-2)+a})^2 \gamma_{b-2}^{(a)} \\ &= \bar{w}_s w_{s(b-1)+a} \left\{ \left(1 - \frac{\bar{w}_s w_{s(b-2)+a}}{w_{s(b-1)+a}} \right) w_{s(b-2)+a} \gamma_{b-2}^{(a)} - \bar{w}_s^2 w_{s(b-3)+a}^2 \gamma_{b-3}^{(a)} \right\} \\ &\geq \bar{w}_s w_{s(b-1)+a} \left(\bar{w}_s w_{s(b-2)+a} \gamma_{b-2}^{(a)} - \bar{w}_s^2 w_{s(b-3)+a}^2 \gamma_{b-3}^{(a)} \right) \\ &= \bar{w}_s w_{s(b-1)+a} \beta_{b-1}^{(a)}. \quad \square \end{aligned}$$

Now, let us consider the kernel of the following form. In the same way as in the above, $k(x_i, y_j)$ is a positive semidefinite kernel over Σ.

$$K(x, y) = \sum_{i=1}^{L} w_i \left[k(x_i, y_i) + \sum_{s=1}^{S} \bar{w}_s \left\{ k(x_{i+s}, y_i) + k(x_i, y_{i+s}) \right\} \right]. \tag{19}$$

Let $b'^{(s)}_{w_1,\dots,w_L}$ be positive numbers such that, if $0 \le \bar{w}_s \le b'^{(s)}_{w_1,\dots,w_L}$, the kernel of Eq. (14) is positive semidefinite. If we have $\sum_{s=1}^{S} \alpha_s = 1$ such that $0 \le \bar{w}_s \le \alpha_s b'^{(s)}_{w_1,\dots,w_L}$, $K_s(x, y)$ defined below is positive semidefinite, and therefore so is $K(x, y) = \sum_{s=1}^{S} K_s(x, y)$.

$$K_s(x, y) = \sum_{i=1}^{L} w_i \left[\alpha_s k(x_i, y_i) + \bar{w}_s \left\{ k(x_{i+s}, y_i) + k(x_i, y_{i+s}) \right\} \right].$$

Thus, we have obtained Corollary 21.

Corollary 21. *If the following inequality holds for $\bar{w}$, the character-base string kernel of Eq.* (19) *is positive semidefinite for an arbitrary positive semidefinite $k(x_i, y_j)$.*

$$\sum_{s=1}^{S} \frac{\bar{w}_s}{b'^{(s)}_{w_1,\dots,w_L}} \le 1.$$

Proof. We have only to take α_s such that $\frac{\bar{w}_s}{b'^{(s)}_{w_1,\dots,w_L}} \le \alpha_s$ and $\sum_{s}^{S} \alpha_s = 1$. □

The sufficient condition by [4], which was also described in 3.1.3 is obtained as a corollary to Corollary 20 and Corollary 21.

Corollary 22. *If $w_1 \le \cdots \le w_L$, the kernel of Eq.* (19) *with $\sum_{s=1}^{S} \bar{w}_s \le \frac{1}{2}$ is positive semidefinite.*

9. Conclusion

The main theorem of this paper (Theorem 9) considerably generalizes the well-known polynomial kernel. While the polynomial kernel only takes advantage of the univariate polynomials with non-negative coefficients to engineer positive semidefinite kernels, our theorem expands the domain of polynomials to multivariate polynomials with positive and negative coefficients. More specifically, the main theorem asserts the positive semidefiniteness of the kernels obtained by applying a polynomial to positive semidefinite underlying kernels, if the polynomial has a positive semidefinite coefficient matrix. Not only does the theorem support the positive semidefiniteness of many kernels known in the literature, but it also enlarges our latitude in engineering positive semidefinite kernels. As evidence for this, by taking advantage of the theorem, we presented extended and corrected conditions for the codon-improved kernel and the weighted degree kernel with shifts to be positive semidefinite.

References

[1] C. Berg, J.P.R. Christensen, R. Ressel, Harmonic Analysis on semigroups, in: Theory of Positive Definite and Related Functions, Springer, 1984.
[2] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, 2000.
[3] C.S. Leslie, E. Eskin, W.S. Noble, The spectrum kernel: A string kernel for SVM protein classification, in: Pacific Symposium on Biocomputing, 2002.
[4] G. Rätsch, S. Sonnenburg, B. Schölkopf, Rase: Recognition of alternatively spliced exons in C.elegans, Bioinformatics 21 (2005) i369–i377.
[5] K. Shin, Position-aware string kernels with weighted shifts and a general framework to apply string kernels to other structured data, in: The 8th International Conference on Intelligent Data Engineering and Automated Learning, 2007.
[6] S. Vishwanathan, A. Smola, R. Vidal, Binet–Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes, International Journal of Computer Vision 73 (2007) 95–119.
[7] L. Wolf, A. Shashua, Learning over sets using kernel principal angles, Journal of Machine Learning Research 4 (2003) 913–931.
[8] O. Yamaguchi, K. Fukui, K. Maeda, Face recognition using temporal image sequence, in: IEEE International Conference on Automatic Face & Gesture Recognition, 1998.
[9] S.K. Zhou, Trace and determinant kernels between matrices, SCR Technical Report, 2004. URL citeseer.ist.psu.edu/662320.html.
[10] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, K.R. Müller, Engineering support vector machine kernels that recognize translation initiation sites, Bioinformatics 16 (9) (2000) 799–807.